DOCUMENT RESUME

ED 165 504 FL 010 057

AUTHOR	Bjurlof, Thomas; Jamieson, Dale
TITLE	Semantics and the Number of English Sentences.
PUB DATE	78
NOTE	17p.
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Comprehension; Deep Structure; *English; Grammar; Language Patterns; *Linguistic Theory; Logic; Nominals; *Semantics; *Sentences; *Sentence Structure; Surface Structure; Syntax
IDENTIFIERS	*Chomsky (Noam)

ABSTRACT

It has long been said that there are an infinite number of English sentences. "This is the cat that caught the rat" is an English sentence. So is "This is the cat that caught the rat that stole the cheese." "This is the cat with white paws that caught the rat that stole the cheese" is unobjectionable as well. Since a clear cutoff point cannot be specified, it is tempting to resort to the three dots. This study proposes that the argument from the lack of a clear cutoff point to infinity is a bad argument; the set of English sentences may be a fuzzy set rather than a standard set. Furthermore, it is argued that the initial question suppresses several quite distinct questions. A theory constructed to account for formal relations between sentences might warrant the positing of infinite semantic structures. A theory constructed to account for human understanding probably would not. The common claim that infinity is necessary if there are to be novel and creative uses of language is found to be entirely without substance. (Author/AMH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED165504

SEMANTICS AND THE NUMBER OF ENGLISH SENTENCES

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Dale Jamieson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

by

Thomas Bjurlöf
Department of Philosophy
University of North Carolina
Chapel Hill, NC 27514

Dale Jamieson
Department of Philosophy
State University of New York
College at Fredonia
Fredonia, NY 14063

FL010057

Semantics And The Number of English Sentences

1. How many English sentences can be understood? Very, very many. Noam Chomsky has said and perhaps would still say indefinitely many or infinitely many (not noticing the difference, of which more later).[1] How many English sentences are there? Zellig Harris has said and perhaps would still say denumerably infinitely many.[2] Many or very many is not enough according to Chomsky and Harris. The difference between very large numbers and infinity is supposed to reflect interesting features of human languages.

2. These views have recently been challenged by Paul Ziff.[3]

> ...there is no important sense in which it is true that there are infinitely many English sentences.[4]

Ziff would agree that there is an unimportant sense in which there are infinitely many English sentences. One can do arithmetic in English. But Ziff thinks that this possibility "is not illuminating with respect to English structure or English grammar or anything like it."[5] Ziff has no quarrel with syntactic components that generate infinitely many syntactic structures. His qualms enter with semantics.

> Although it is child's play to characterize powerful syntactic devices capable of generating an infinity of sentences there's no reason to believe that the semantics of English (or of any natural language) is of a consonant or even comparable character.[6]

And he gones on.

> Inevitably one will produce nonsensical sentences or sentences that are utterly incomprehensible to speakers of the language...[7]

3. Jonathan Bennett has responded to Ziff's challenge. In his Linguistic Behaviour Bennett replies to Ziff with the following.

> But the question 'How many English sentences are there?' does not mean 'How many sentences are there which would be understood by speakers of English?' and no one would give the answer 'infinitely many' to the latter question...The notion of what has meaning is admittedly founded upon the notion of what is understood, but the two are not co-extensive.[8]

Bennett supposes that the notion "what has meaning" and the notion "what is understood" have extensions. Thus he supposes that there is some class whose members are "what has meaning" and another class whose members are "what is understood." Bennett also talks about "the class of 'English sentences'".[9] But these predicates are not well-defined. Hence there is no such thing as "the class of 'English sentences'", nor the extensions of the notions "what has meaning" and "what is understood."[10] But if we suppose that it does make sense to talk about extensions in this context, Bennett is certainly correct: There are English sentences that will never be understood by English-speakers. Some since they will never be uttered, some since they are too complex, some for other reasons. Bennett assumes a stronger thesis

however. He assumes that there are English sentences that could not be understood by English-speakers.

4. If Ziff's position were that being a sentence (or having meaning) is co-extensive with being understood, then Bennett would have an easy victory and Ziff's paper could largely be ignored. Things are not that simple, however.

Let us look at an example. Ordinarly we have no difficulty in understanding the following sentences:

(a) My uncle is a linguist.

(b) My father's brother is a linguist.

(a) and (b) have (roughly) the same meaning.[11] (Notice that one might very well know the meaning of a sentence or that two sentences are synonymous without having understood the sentences. Someone might know that the sentence written on a piece of paper in his left pocket has (roughly) the same meaning as the sentence written on a piece of paper in his right pocket. Someone might have told him so. Nevertheless he might not have understood either sentence. He might never have looked at them.) Suppose that from tomorrow on English speakers cannot make heads or tails of subordinate groups of two nouns if the nouns can be reversed. Examples of such pairs would be 'mother's daughter', 'father's brother' and 'a circle under a triangle'. Would Ziff say that since (b) cannot be understood by English speakers, although (a) presents no problems, (b) is not an

an English sentence? We hope not. Failure to understand (b) need not be attributed to the semantics of (b), but should most likely be attributed to neurological factors. After all, (a) and (b) have (roughly) the same meaning. (Roman Jakobson discusses a type of aphasia that would explain our case.[12]) What Ziff would say is that failure to understand a candidate for an English sentence may disqualify it as an English sentence if the failure can be attributed to semantic factors.[13] He does not say that a candidate for an English sentence must be disqualified if the failure can be attributed to neurological factors.

5. One has to be careful in attributing the failure to understand a sentence to semantic factors. Some cases are fairly straight forward. The failure of someone under some conditions to understand

(c) 7 is green.

could be attributed to the fact that a number is not the kind of thing that is colored; and this might be thought of as a semantic factor. Notice that we don't claim that anyone would fail to understand an utterance of (c) under any conditions whatever, just that there are conditions under which someone would fail to understand (c) due to semantic factors.

Why do we attribute this failure of understanding to semantic factors? Because we take (c) to have the syntactic structure NP + VP, and hence to be syntactically well-formed.

Thus, this failure of understanding must be due to semantic factors. Contrast (c) with 'Ball dolphin mop cathedral.'; a string that has the syntactic structure NP + NP + NP + NP, and thus is not syntactically well-formed. (Of course schemes have been devised that count all deviance as syntactic. Similarly schemes have been devised that count all deviance as semantic. Neither seem particularly well-motivated.[14])

Further complications are presented by other cases. For example

(d) This sentence is in English.

If someone cannot understand an utterance of (d) because he cannot figure out that 'this' in (d) refers to (d), the failure to understand (d) cannot be attributed to semantic factors, namely, the reference of 'this' in (d) on this occasion. Rather it can be attributed to the failure to appreciate the relevant semantic factors. But don't confuse the last case with the following:

(e) This sentence is false.

One could argue that, if conceived self-referentially, (e) could not be understood by any English speaker. (e) is, of course, a version of the well-known "liar paradox." If we assume it to be true it follows that it is false and vice versa. No one knows what it would be like for (e) to be true, or what it would be like for it to be false. We will not pursue this line of argument.[15]

6. Many linguists claim that that there are infinitely many English sentences. It is often unclear what question this is an answer to. How many English sentences are there, of course; but as we shall see, what is asked by this question is far from clear. Here is an example:

> It is astonishing to find that even this truism [the implicit ability to understand indefinitely many sentences] has recently been challenged. See Dixon (1963). However, it seems that when Dixon denies that a language has infinitely many sentences, he is using the term "infinite" in some special and rather obscure sense. Thus on the same page (p. 83) on which he objects to the assertion "that there are an infinite number of sentences in a language" he states that "we are clearly unable to say that there is any definite number, N, such that no sentence contains more than N clauses" (that is, he states that the language is infinite). Either this is a blatant self-contradiction, or else he has some new sense of word "infinite" in mind.[16]

But surely the mistake is Chomsky's, and Dixon is not contradicting himself. On Chomsky's view, that there is no definite number of English sentences is equivalent to the claim that there are infinitely many English sentences. Why not, not infinitely many but no definite number? In standard set theory[17] if a set has no definite finite number of elements it has infinitely many elements. And all finite sets have a definite number of elements. But these are standard sets where membership is well-defined. Something is either a member or it is not. If we take another approach and think of the set of English sentences as a fuzzy set indefinite does not

imply infinite. In fuzzy set theory membership is graded. An element is assigned a number in the inclusive interval [0,1]. The number assigned to an element with respect to a set is the degree to which that element belongs to the set. It does not make sense to ask how many members a fuzzy set has since membership in a fuzzy set is not an either/or question. There is, for example, no answer to the question: How many tall men are there? Some people are taller than others. Some are clearly members of the (fuzzy) set of tall men. Some not so clearly. Some clearly not. But how many tall men are there? No answer. And it does not follow that there are infinitely many tall men.[18]

Maybe Chomsky thought of this: 'This is the cat that caught the rat.' is an English sentence. So is "This is the cat that caught the rat that stole the cheese.'. 'This is the cat with white paws that caught the rat that stole the cheese.' is also an English sentence. And it seems that one can go on like this. For how long? There is no definite cutoff point Chomsky claims, and he is right. Does it follow that one can go on forever? That is, have we shown that there is unbounded recursion in English? Of course not. To show this we need a proof of unbounded recursion and not just a proof that there is no definite cutoff point. Chomsky does not provide such a proof. That there is no definite cutoff point does not create any problems. Maybe we are dealing with fuzzy sets, and then we do not need definite cutoff points.

7. How many English sentences are there? Is the question how many English sentences can be understood? Or how many syntactically well-formed English sentences are there? Or how many syntactic structures a syntactic theory of English should generate? Or how many semantically non-deviant English sentences are there? Or how many English sentences a semantic theory of English should account for? Or how many acceptable English sentences are there?

We will not quarrel with an infinite syntactic component. Maybe infinity allows us to simplify syntactic theory. Maybe it makes it less complicated. Ziff's rejection of infinity is based on semantics, and we will attend to semantic structures in the following.

8. The question is: How many English sentences should a semantic theory of English account for? But a prior question: What should a semantic theory do? Assign meaning to syntactic structures; account for sameness of meaning between sentences, phrases, words and morphemes; account for paraphrase; characterize logical consequence and semantic entailment. These are commonly cited requirements. If we start with an infinite syntactic component and adopt these requirements (or at least some of them, and no others), then a semantic theory should account for an infinite number of English sentences. This is the picture exhibited by philosophers and logicians such as Richard Montague and Donald Davidson. A language is thought of as an abstract structure, not as a means of communication

among humans. But there is a problem. In most formal theories the syntax is constructed so that it fits the semantics, and syntactic and semantic well-formedness coincide. Linguists have often made a distinction between sentences which are syntactically deviant and those which are semantically deviant, and these notions do not coincide. Sentence (c) on page 4 is perhaps syntactically well-formed. That is, there might be a reasonable syntactic theory of English such that (c) does not violate any rule in this theory. But it is semantically deviant. There are difficulties in drawing a line between syntactic and semantic deviance.[20] The distinction is clear enough, however, to allow one to point to sentences that do not violate any syntactic rule but do violate semantic rules. In other words: Not every syntactically well-formed sentence is semantically well-formed. This is enough to show that an infinite syntactic component does not imply an infinite semantic component. We need non-syntactic arguments for this conclusion. One such argument could be that infinity in semantics allows for simpler semantic theories. Bennett argues in this way in Linguistic Behavior.[21] Given the primitive state of semantic theory this argument is difficult to evaluate, so we will remain agnostic on this point.

9. Something is missing from the picture exhibited by Montague and Davidson. Sometimes more is wanted from a semantic theory than that it account for formal semantic relations. Some have wanted semantic theory to aid in the

understanding of what it is to understand a sentence (phrase, word, morpheme). This concern with understanding is a characteristic of Ziff's work in linguistics and semantics.[22] It serves to distinguish Ziff's work from that of Montague and Davidson.

Is there anything infinite about understanding? In a trivial sense the answer is no. People's lives are finite, attention-spans are short, memories are limited.

Linguists in the Chomsky-Katz mold also claim to be concerned with understanding:

> We may thus regard the development of a model of the semantic component as taking up the explanation of a speaker's ability to produce and understand indefinitely many new sentences.[23]

But they are unimpressed with short attention-spans and limited memories:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance.[24]

These linguists explicitly hold that there is an infinite number of English senteces. How can the concern with understanding be reconciled with this claim? In the following way: Linguistics is not about "cheap empirical facts"; it is about the behavior of ideal speaker-hearers.

Two lines of assault can be mounted against this position.

One could disagree with the goals of linguistic theory as conceived by Chomsky and Katz. This we shall not do; it has been well done elsewhere.[25] Instead we will take another line of assault. Let us require of a semantic theory that it account for the strategies employed by speaker-hearers in understanding sentences.

What are strategies? A strategy would not be so-called if it did not aid in achieving what it is a strategy for. Strategies are like algorithms; both are procedures for solving problems. However, not all algorithms are like strategies; strategies are efficient procedures for solving problems. There are problems for which algorithms exist but require so much time for their implementation that in practice they cannot be employed. (The only known algorithms that are solutions to the problem of the traveling salesman's tour is an example.[26]) The strategies employed by human speaker-hearers cannot be of this kind. If they were, the rapidity of human understanding would be left unexplained. But this argument for the efficiency of the strategies employed by human speaker-hearers is not an argument for the finiteness of such strategies. It could be the case that there are strategies like the following:

a means

aa means what a means

aaa means what a means

...

That is, every sequence of a's means what the initial a means. This strategy would be efficient and would account for an infinite number of sentences. Here the problem is not with understanding but with production. Human speaker-hearers do not and could not produce such sentences (at least past some unspecified number of iterations). Since human speaker-hearers do not and could not produce such sentences, there is no reason to believe that such strategies are employed in understanding.

What Ziff says is right: It is not mere chance that only a finite number of sentences are uttered and understood; it could not be otherwise. The strategies employed by human speaker-hearers in understanding are finite, however ideal you want these humans to be; and what humans operate on is finite as well.

An infinity of sentences is a lot. One can go on and on and on and on and on and on and not reach infinity. How many English sentences are there? Probably not as many as there are electrons in the universe. Probably not 10^{79}.

But you may not be swayed. What about novelty? What about creativity? What about the so-called "rationalist" tradition in linguistics? One bad argument that is repeated in the linguistic literature is that infinity is required if we are to say new and interesting things. But as almost everyone knows, only a finite number of English sentences have been and will be uttered. Finnegan's Wake, 77 Dream Songs and all of Shakespeare and Strindberg were written from finite resources. But again one

might say that none of these works could have been written were there not an infinite number of unuttered sentences. One would have thought that linguistics is an empirical science. Now we discover that it haunts the realm of "unactualized possibles." Linguists who make such claims devour themselves.[27]

NOTES

[1]N. Chomsky, Aspects of the Theory of Syntax, MIT Press, 1965, pp. 3-62.

[2]Z. Harris, Mathematical Structures of Language, Interscience Publishers, 1968, pp. 9-11.

[3]P. Ziff, "The Number of English Sentences," Foundations of Language 1974, pp. 519-532.

[4]Ibid., p. 519.

[5]Ibid., p. 521.

[6]Ibid., pp. 529-530.

[7]Ibid., p. 530.

[8]J. Bennett, Linguistic Behaviour, Cambridge University Press, 1976 pp. 279-280.

[9]Ibid., p. 279.

[10]On this point see Dana Scott's excellent essay, "Background to Formalization," in Truth, Syntax and Modality, H. LeBlanc (ed.), North Holland, 1973, p. 244-273.

[11]There is a complication that we shall ignore: An uncle can be a mother's brother or an aunt's husband rather than a father's brother.

[12]In R. Jakobson, Studies on Child Language and Aphasia, Mouton 1971, pp. 75-94.

[13]In fact he has said this in conversation. It is difficult to find this view stated explicitly in "The Number of English Sentences," however.

[14]See for example: J. McCawley, "Concerning the Base Component in a Transformational Grammar", Foundations of Language 1968, pp. 243-269; and K. Antley, "McCawley's Theory of Selectional Restriction, Foundations of Language 1974, pp. 257-272.

[15]For this view, see P. Ziff, Semantic Analysis, Cornell University Press, 1960, pp. 136-138.

[16] N. Chomsky, op.cit., footnote 9, p. 198. The inserted clause is from p. 15. The reference to Dixon is to his Linguistic Science and Logic, Mouton & Co., 1963.

[17] An elementary introduction to set theory can be found in R. Stoll, Set Theory and Logic, W. H. Freeman and Company, 1963.

[18] Fuzzy set theory has been developed in great detail. The locus classicus is Zadeh, "Fuzzy Sets," Information and Control 1965, pp. 338-353.

[19] See, for example, the papers by Montague and Davidson in Semantics of Natural Language, D. Davidson and G. Harman (eds.), Reidel, 1972.

[20] For discussions of some of these difficulties, see: U. Weinrich, "Explorations in Semantic Theory," in Current Trends in Linguistics Volume III, T. Sebeok (ed.), Mouton, 1966; P. Ziff, Philosophic Turnings, Cornell University Press, 1966, pp. 119-133; G. Lakoff, "On Generative Semantics," in Semantics, D. Steinberg and L. Jakobovits (eds.), Cambridge University Press, 1971, pp. 232-296.

[21] Op. cit., pp. 277-278.

[22] See, for example, Understanding Understanding, Cornell University Press, 1972.

[23] J. Katz, Philosophy of Language, Harper & Row, 1966, p. 151.

[24] N. Chomsky, op.cit., p. 3.

[25] See, for example, D. Hymes, "Toward Linguistic Competence, Working Papers in Sociolinguistics 16, University of Texas at Austin, Department of Anthropology; G. Harman, "Psychological Aspects of the Theory of Syntax," Journal of Philosophy 1967, pp. 75-87; B. Derwing, Transformational Grammar as a Theory of Language Acquisition, Cambridge University Press, 1973, pp. 259-296.

[26] See H. Lewis and C. Papadimitriou, "The Efficiency of Algorithms," Scientific American 1978, pp. 96-109.

[27] An earlier version of this paper was presented at a conference on "Productivity, Novelty, and Creativity in Language" at the University of Louisville. We wish to thank Jonathan Bennett and Paul Ziff for valuable correspondence and conversation concerning these topics. Thanks also to R. J. Haack and Richard Sharvy.